

Real Space Refinement of Crystal Structures with Canonical Distributions of Electrons

Simon W. Ginzinger,¹ Markus Gruber,¹ Hans Brandstetter,² and Manfred J. Sippl^{1,*}

¹Division of Molecular Biology, Department of Bioinformatics, University of Salzburg, Hellbrunnerstraße 34, 5020 Salzburg, Austria

²Division of Molecular Biology, Department of Structural Biology, University of Salzburg, Billrothstraße 11, 5020 Salzburg, Austria

*Correspondence: sippl@came.sbg.ac.at

DOI 10.1016/j.str.2011.10.011

Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

SUMMARY

Recurring groups of atoms in molecules are surrounded by specific canonical distributions of electrons. Deviations from these distributions reveal unrealistic molecular geometries. Here, we show how canonical electron densities can be combined with classical electron densities derived from X-ray diffraction experiments to drive the real space refinement of crystal structures. The refinement process generally yields superior molecular models with reduced excess electron densities and improved stereochemistry without compromising the agreement between molecular models and experimental data.

INTRODUCTION

X-ray crystallography reveals the three-dimensional structures of biological macromolecules with atomic details of functional sites and disease-relevant aberrations (Wimberly et al., 2000; Ban et al., 2000; Schlutzen et al., 2000; Cramer et al., 2001; Dutzler et al., 2002; Kanamaru et al., 2002). The essential milestone in the determination of crystal structures is the compilation of electron densities from measured amplitudes and their interpretation in terms of atomic coordinates. The reliability of molecular structures obtained in this way is generally limited by the resolution of the respective diffraction pattern and is further compromised by the crystallographic phase problem. Moreover, for large protein complexes and membrane proteins, it is difficult to grow crystals of sufficient quality, severely limiting the achievable resolution. All these problems combine to impede the construction and interpretation of three-dimensional electron densities.

With the goal to facilitate the determination of molecular structures from electron densities, the crystallographic community has invented a number of techniques including methods that exploit noncrystallographic symmetry (Bricogne, 1976), the flatness of solvent regions (Wang, 1985), and standardized angle and bond parameters (Engl and Huber, 1991) (for a summary see Rupp, 2009). Moreover, a number of diagnostic tools are available that flag inconsistencies in torsion angles (Weichenberger and Sippl, 2007; Word et al., 1999a; Ramachandran et al., 1963) and atoms in unrealistically close contact (Chen et al., 2010; Hooft et al., 1996). However, these tools rarely provide the means for the correction of detected inconsistencies.

Many of the problems encountered in the refinement of crystal structures originate from inconsistencies in the description of the allowed geometries of densely packed groups of atoms. We have recently observed that recurring configurations of atoms in molecules are surrounded by characteristic distributions of electrons (Ginzinger et al., 2010). These distributions are canonical in the sense that they are largely invariant and independent of the specific molecule in which they reside. It follows immediately that realistic atomic models of molecular structures have to satisfy these distributions.

Here, we apply the canonical distributions of electrons to the real space refinement of crystal structures. The procedure, called canonical density expansion (CDE), consists of two steps. We first scan crystal structures for regions that have excess electron densities relative to the expected canonical densities. Second, we remove these excess densities in subsequent refinement cycles using a combination of canonical and experimental densities as a target function. In what follows we briefly review the compilation of canonical densities. We then provide several examples of CDE refinement and discuss the results obtained. Technical details are deferred to the [Experimental Procedures](#) section.

RESULTS AND DISCUSSION

Generally, molecules can be completely dissected into small quasi-rigid molecular fragments composed of a few atoms, like hydroxyl and carbonyl groups, or aromatic rings. In particular the amino acids in protein molecules may be dissected into fragments containing no or negligible internal degrees of freedom, i.e., atomic configurations devoid of rotatable bonds, like carboxylates (Asp, Glu), ring systems (Phe, Trp, His; Pro), guanidino groups (Arg), and carbonyl and amid groups of peptide bonds. The corresponding canonical distributions of electrons can be obtained by averaging over ensembles of such fragments derived from a large number of molecules whose electron density distributions are known from high-resolution X-ray analysis (Ginzinger et al., 2010). In our method of CDE, we take advantage of the spatial information contained in canonical distributions of electrons.

We illustrate the principles of CDE using the tricorn-interacting factor F1 from the archaeon *Thermoplasma acidophilum*, Protein Data Bank (PDB) entry 1xqy, as an example. In that particular case inhibitor soaking had cracked the native F1 crystals, impairing the quality and resolution of the diffraction pattern (Goettig et al., 2005). Despite these difficulties the structure of the F1 complex was solved at 3.2 Å resolution using a related 2.0 Å

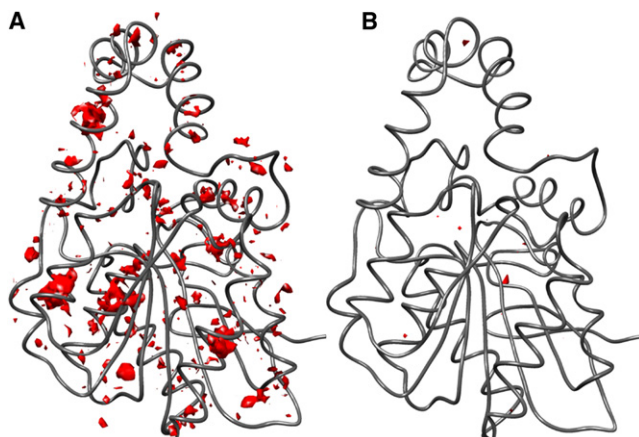


Figure 1. Difference Density Map ($\rho_{\text{canon}} - \rho_{\text{calc}}$) Reveals Regions of Excess Electron Density in Protein Structures

(A) The deposited model of 1xqy contains many regions of large excess electron density.

(B) These regions largely disappear after refinement by CDE. Regions of excess electron density are shown in red, contoured at $-0.2 \text{ e}^-/\text{\AA}^3$.

F1 structure for phasing. The high R_{free} value of 36% of the structural model was ascribed to the multiply twinned diffraction pattern of the soaked crystals rather than to errors in the phases. However, when we re-refine the published F1 structure by CDE, we obtain a significantly improved structural model, as we demonstrate below.

The canonical density ρ_{canon} of 1xqy is constructed by mapping the molecular fragments contained in the precompiled library of canonical densities onto the local coordinate frames of the corresponding molecular fragments found in 1xqy. The complete density ρ_{canon} is obtained by summing over the densities of the individual fragments where regions of overlapping volumes are properly averaged. Subtracting the model density ρ_{calc} then yields the difference density $\rho_{\text{canon}} - \rho_{\text{calc}}$ in familiar units of electrons per cubic Angstrom, $\text{e}^-/\text{\AA}^3$. Negative values in the difference density correspond to regions of electron

excess in the model density relative to the expected canonical density (Figure 1).

To correct excess electron densities, we combine the crystallographically observed electron density ρ_{obs} with the difference density to obtain the hybrid electron density distribution,

$$\rho_{\text{hybrid}} = \rho_{\text{obs}} + \rho_{\text{canon}} - \rho_{\text{calc}} \quad (1)$$

Because the ρ_{hybrid} density is technically equivalent to a $2F_{\text{obs}} - F_{\text{calc}}$ map, an appropriately formatted hybrid density is compatible with the major crystallographic programs (Collaborative Computational Project, 1994; Kleywegt and Jones, 1996; Turk, 2000; Pettersen et al., 2004; Emsley et al., 2010). Hence, ρ_{hybrid} maps can be used immediately in real space refinement, either manually, or by applying some automated refinement protocol. To demonstrate the point of principle, we choose a simple automated strategy using the Coot program (Emsley et al., 2010). The hybrid map is loaded into Coot, and the highest excess density peak in ρ_{hybrid} is located. The corresponding excess density is then minimized by moving the atomic coordinates contributing to this peak. The minimization is terminated when there is no further reduction in excess density. The procedure is repeated for the second-highest excess density peak and so on, where we cycle through the molecule as long as improvements are observed. Applying this protocol of iterated CDE to 1xqy, we obtain a considerably improved structural model as indicated by a 63% decrease in the total volume of excess electron density accompanied by a decrease in R_{free} of 2.1% (Figure 1; Table 1).

The root-mean-square error of optimal superposition, R_s , between 1xqy and the re-refined model is 0.187 \AA , indicating that the refinement did not result in any large-scale movements. However, this is an average over many distances. There are in fact appreciable rearrangements in parts of the molecule. For example in the region consisting of residues 183–187, we observe a deviation of $R_s = 0.356 \text{ \AA}$. When we compare the refined 1xqy coordinates to 1mtz, a 1.80 \AA high-resolution mate of 1xqy, we find that R_s remains practically constant (0.681 \AA in the starting model and 0.677 in the re-refined model). Hence, overall, the

Table 1. CDE Refinement of Crystal Structures

Code	Resol	Size	V_E	V_E^{CDE}	$\Delta V_E(\%)$	C	C^{CDE}	$\Delta C(\%)$	$R_{\text{free}}(\%)$	$R_{\text{free}}^{\text{CDE}}(\%)$	$\Delta R_{\text{free}}(\%)$
1xqy	3.20	294	209.38	77.25	63.10	39,375	29,447	25.21	35.57	33.43	2.14
1orw	2.84	2,912	668.12	441.88	33.86	242,757	219,326	9.65	24.00	24.28	-0.28
1z1w	2.70	780	164.12	111.25	32.21	73,079	66,194	9.42	29.20	29.04	0.16
1jh1	2.70	158	22.75	9.00	60.43	10,960	7,868	28.21	27.54	27.27	0.27
1klj	2.44	304	72.50	23.50	67.58	23,219	18,333	21.04	28.83	28.80	0.03
3gdg	2.30	1,068	212.00	112.00	47.16	80,363	64,043	20.30	29.11	29.41	-0.30
2wpk	2.21	297	73.00	57.88	20.71	16,645	16,387	1.55	24.95	24.82	0.13
1k32	2.00	6,138	1,125.12	893.62	20.57	420,657	402,786	4.24	29.76	29.75	0.01
1xro	1.80	290	46.62	38.62	17.16	20,887	19,693	5.71	25.79	25.93	-0.14
1kli	1.69	315	54.25	38.62	28.81	17,689	16,859	4.69	27.17	26.44	0.73

The number of atoms in close contact is calculated using the Probe program (Word et al., 1999a). Here, C and C^{CDE} are the number of overlaps in the deposited model and CDE-refined model, respectively. ΔC is the corresponding change in the number of overlaps. R_{free} values are calculated from the structure factors and coordinates deposited with the PDB. $R_{\text{free}}^{\text{CDE}}$ values are calculated from the deposited structure factors and the re-refined model obtained after iterative CDE. ΔR_{free} is the change in the crystallographic R_{free} value. Code, PDB accession code; Resol, resolution in (\AA) of the corresponding crystal structure; Size, number of residues in the asymmetric unit; V_E , total volume of excess electron density for the deposited model; V_E^{CDE} , total volume of excess electron density for the CDE-refined model; ΔV_E , corresponding reduction in the total volume of excess electron density.

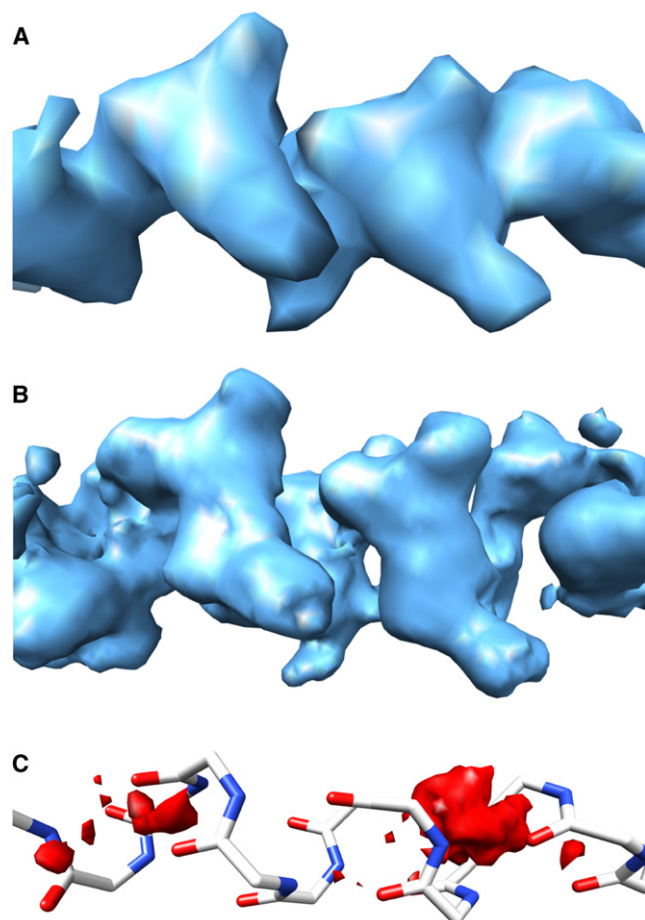


Figure 2. Electron Density Maps Used in CDE

The synthesis of hybrid densities $\rho_{\text{hybrid}} = \rho_{\text{obs}} + \rho_{\text{canon}} - \rho_{\text{calc}}$ is visualized using the helical segment, amino acid residues 281–293 of 1xqy.

(A) Experimental density ρ_{obs} contoured at $0.15 \text{ e}^-/\text{\AA}^3$.

(B) Hybrid density ρ_{hybrid} contoured at $0.15 \text{ e}^-/\text{\AA}^3$.

(C) Molecular model and the corresponding density difference $\rho_{\text{canon}} - \rho_{\text{calc}}$ contoured at $-0.2 \text{ e}^-/\text{\AA}^3$.

CDE refinement of 1xqy results in local rearrangements but does not change the basic conformation of the molecule.

With decreasing resolution the interpretation of electron densities becomes increasingly difficult. For example at a resolution worse than 3 \AA , local electron density peaks merge, and maxima appear at locations that do not correspond to atomic positions. Illustrative examples are phenyl and tyrosine side chains where peaks appear at the centers of ring structures. A similar phenomenon is observed at medium and low resolution in the carbonyl collapse of helices where the electron density maxima appear along the helix axes. During crystallographic refinement the high electron density along the helix axis attracts atoms of the protein backbone resulting in distorted peptide bonds with carbonyl oxygens pointing toward the helix center. As a consequence ρ_{calc} is in good agreement with the poorly resolved electron density, but the resulting atomic model is unrealistic and needs to be corrected by noncrystallographic restraints. The corresponding inconsistencies in helix geometry are easy to spot, but their correction is generally cumbersome because

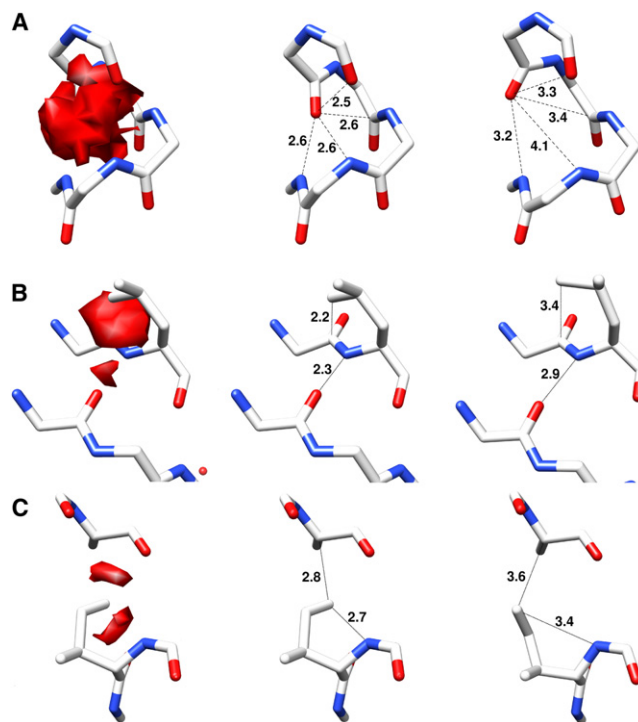


Figure 3. Examples of CDE-Refined Structures

(A) Electron density excess (left) in the region of residues A282–A287 of 1xqy, atomic distances (center), and distances of the CDE-refined model (right).

(B) Backbone to side-chain clash (residues D170 and D171) and jolted hydrogen bond (residues D171 and D253) in the structure of the Sigma-54 transport activator (3n70, resolution 2.80 \AA) producing a small but intense region of electron excess (left), the corresponding atomic distances (center), and the CDE-refined model (right).

(C) Excess electron density (left), distances (center), and CDE-refined structure (right) of residues A1136 and A1141 of human ubiquitin F box ligase complexed with S phase kinase-associated protein 1 (3l2o, resolution 2.80 \AA).

the subsequent refinement tends to repeat the same error. In contrast, CDE refinement against the hybrid electron density ρ_{hybrid} enforces realistic molecular geometries (Figures 2 and 3).

To investigate the general applicability of CDE in X-ray analysis, we initially re-refined ten crystal structures previously solved and published by us (Table 1). In all these cases automated iterated CDE reduces the excess electron density that is generally correlated with reductions in atomic clashes as reported by the program Probe (Word et al., 1999b). These findings are confirmed on a larger scale when we apply CDE to 128 crystal structures found in a recent weekly release of the PDB (Figure 4). In many cases the excess electron density is reduced substantially, which is generally accompanied by reductions in the number of atomic clashes. It follows that a large number of previously solved structures can be substantially re-refined by CDE, in particular in the medium- and low-resolution range, leading to superior molecular models. To conclude, we emphasize that the difference density $\rho_{\text{canon}} - \rho_{\text{calc}}$ is independent of experimental densities. Therefore, canonical distributions of electrons can be used in the refinement of structural models obtained from NMR, cryo-EM, and other imaging techniques, as well as structure prediction and molecular modeling. Difference

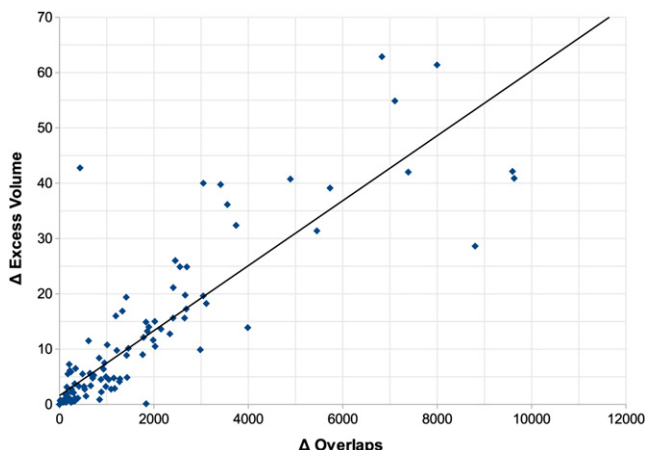


Figure 4. CDE Refinement of Structures Released by the PDB on July 27, 2011

CDE refinement is applied to all 128 X-ray structures of the respective release. The difference in electron excess volume (Δ Excess Volume) is plotted against the number of overlap dots calculated using Probe (Word et al., 1999a) (Δ Overlaps). The linear regression (black line) has the parameters $y = 0.006x + 1.595$ and a correlation of $R^2 = 0.756$.

density maps for structural models of any origin can be computed and downloaded from <http://canden.services.came.sbg.ac.at>

EXPERIMENTAL PROCEDURES

The canonical densities used here are obtained from known protein structures (Berman et al., 2000) with resolution better than 1.5 Å, where multiply solved and homologous proteins are removed using the COPS classification system (Suhre et al., 2009) with a threshold of 90% relative structural similarity (Sippl, 2008; Sippl and Wiederstein, 2008). The procedure yields an unbiased set of 2,383 unique and mutually unrelated crystal structures containing a large number of rigid molecular fragments. The electron density around a particular fragment is calculated using the five Gaussian approximation (Cromer and Waber, 1974) with a constant B factor of 20 Å² for all atoms. This density is mapped on a cubic grid embracing the atoms of the fragment. To sample the electron density around a particular fragment type, we first determine the enclosing surface, which is defined by a constant electron density. Because our goal is to sample the environment around the fragments, we choose a low-density cutoff of 0.1 e[−]/Å³. This ensures that the volumes of fragments that are close in space overlap, which is necessary in order to sample their shared density. The individual fragments of a given type are then superimposed relative to a common reference frame together with their respective molecular environments, which may consist of other amino acids, nucleic acids, substrates, ions, and other solvent molecules. The canonical electron density for a particular fragment type is then computed as the ensemble average over the corresponding grids.

The canonical density, ρ_{canon} , of a particular molecular model is obtained by superimposing the canonical densities of the fragments onto the corresponding atoms of the model. In regions where the volumes of two individual fragments overlap, the sum of the canonical density is averaged. The difference density, $\rho_{\text{canon}} - \rho_{\text{calc}}$, then shows where the model density deviates from the canonical density. The hybrid density, ρ_{hybrid} , is obtained from the combination of ρ_{obs} and $\rho_{\text{canon}} - \rho_{\text{calc}}$ as discussed above. The single-most important quantity is the difference density that shows where a structural model has inconsistencies. The hybrid density is a device for refinement. In manual refinement this is straightforward to apply. However, in automatic refinement the results depend on the refinement program and the specific protocol. Here, we used the Coot program (Emsley et al., 2010) in an automated manner as described in the previous section.

ACKNOWLEDGMENTS

We thank Bernhard Rupp for his advice on several aspects of crystal structure refinement and Ana Caballero-Herrera and Francisco Melo for their critical comments on the manuscript. This work was supported by Austrian Science Fund (FWF): P21294-B12.

Received: March 24, 2011

Revised: September 28, 2011

Accepted: October 10, 2011

Published: December 6, 2011

REFERENCES

- Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* 289, 905–920.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242.
- Bricogne, G. (1976). Methods and programs for direct-space exploitation of geometric redundancies. *Acta Crystallogr. A* 32, 832–847.
- Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 66, 12–21.
- Collaborative Computational Project, Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 50, 760–763.
- Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* 292, 1863–1876.
- Cromer, D.T., and Waber, J.T. (1974). Table 2.2 B. In *International Tables for X-ray Crystallography, Volume IV* (Birmingham: Kynoch Press), pp. 99–101.
- Dutzler, R., Campbell, E.B., Cadene, M., Chait, B.T., and MacKinnon, R. (2002). X-ray structure of a ClC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity. *Nature* 415, 287–294.
- Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486–501.
- Engh, R.A., and Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr. A* 47, 392–400.
- Ginzinger, S.W., Weichenberger, C.X., and Sippl, M.J. (2010). Detection of unrealistic molecular environments in protein structures based on expected electron densities. *J. Biomol. NMR* 47, 33–40.
- Goettig, P., Brandstetter, H., Groll, M., Göhring, W., Konarev, P.V., Svergun, D.I., Huber, R., and Kim, J.-S. (2005). X-ray snapshots of peptide processing in mutants of tricorn-interacting factor F1 from *Thermoplasma acidophilum*. *J. Biol. Chem.* 280, 33387–33396.
- Hooft, R.W., Vriend, G., Sander, C., and Abola, E.E. (1996). Errors in protein structures. *Nature* 381, 272. <http://dx.doi.org/10.1038/381272a0>.
- Kanamaru, S., Leiman, P.G., Kostyuchenko, V.A., Chipman, P.R., Mesyanzhinov, V.V., Arisaka, F., and Rossmann, M.G. (2002). Structure of the cell-puncturing device of bacteriophage T4. *Nature* 415, 553–557.
- Kleywegt, G.J., and Jones, T.A. (1996). xdlMAPMAN and xdlDATAMAN—programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets. *Acta Crystallogr. D Biol. Crystallogr.* 52, 826–828.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612.
- Ramachandran, G.N., Ramakrishnan, C., and Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* 7, 95–99.
- Rupp, B. (2009). *Biomolecular Crystallography: Principles Practice, and Applications to Structural Biology* (New York: Garland Science).

- Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F., and Yonath, A. (2000). Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell* 102, 615–623.
- Sippl, M.J. (2008). On distance and similarity in fold space. *Bioinformatics* 24, 872–873.
- Sippl, M.J., and Wiederstein, M. (2008). A note on difficult structure alignment problems. *Bioinformatics* 24, 426–427.
- Suhrer, S.J., Wiederstein, M., Gruber, M., and Sippl, M.J. (2009). COPS—a novel workbench for explorations in fold space. *Nucleic Acids Res.* 37 (Web Server issue), W539–W544.
- Turk, D., 2000. MAIN 96: an interactive software for density modifications, model building, structure refinement and analysis. *Proceedings from the 1996 meeting of the International Union of Crystallography Macromolecular Computing School*, eds.
- Wang, B.C. (1985). Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol.* 115, 90–112.
- Weichenberger, C.X., and Sippl, M.J. (2007). NQ-Flipper: recognition and correction of erroneous asparagine and glutamine side-chain rotamers in protein structures. *Nucleic Acids Res.* 35 (Web Server issue), W403–W406. <http://dx.doi.org/10.1093/nar/gkm263>.
- Wimberly, B.T., Brodersen, D.E., Clemons, W.M., Jr., Morgan-Warren, R.J., Carter, A.P., Vornrhein, C., Hartsch, T., and Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. *Nature* 407, 327–339.
- Word, J.M., Lovell, S.C., Richardson, J.S., and Richardson, D.C. (1999a). Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* 285, 1735–1747. <http://dx.doi.org/10.1006/jmbi.1998.2401>.
- Word, J.M., Lovell, S.C., LaBean, T.H., Taylor, H.C., Zalis, M.E., Presley, B.K., Richardson, J.S., and Richardson, D.C. (1999b). Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. *J. Mol. Biol.* 285, 1711–1733. <http://dx.doi.org/10.1006/jmbi.1998.2400>.